Low Rank Approximations

Note Title

LECTURE 16

Recall Outer product in Lecture 3.

Let $u \in \mathbb{R}^m = \mathbb{R}^{m\times 1}$,
$v \in \mathbb{R}^n = \mathbb{R}^{n\times 1}$.

Then, the outer product between u and v is:

$$u v^T = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} u_1 v_1 & \cdots & u_1 v_n \\ \vdots & & \vdots \\ u_m v_1 & \cdots & u_m v_n \end{bmatrix} \in \mathbb{R}^{m\times n}$$

This matrix has rank 1 because
$u v^T = [v_1 u, \cdots, v_n u]$
i.e., each column is just a scalar multiple of the same vector u.

Now SVD can be viewed as a sum of rank 1 matrices:

Thm $\quad A = \sum_{j=1}^{r} \sigma_j u_j v_j^T, \quad r = \text{rank}(A)$

(Proof) just obvious!

$$[u_1 \cdots u_m] \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ 0 & & & 0 \ddots 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}$$

///

Among all possible $m \times n$ matrices of rank k $(k \le r)$,

$\sum_{j=1}^{k} \sigma_j u_j v_j^T$ is the best approximation of A in the following sense:

Thm For any k with $0 \le k \le r$, let
$$A_k := \sum_{j=1}^{k} \sigma_j u_j v_j^T$$
If $k = p = \min(m,n)$, then define $\sigma_{k+1} = 0$. Then,
$$\|A - A_k\|_2 = \inf_{\substack{B \in \mathbb{R}^{m\times n} \\ \mathrm{rank}(B) \le k}} \|A - B\|_2 = \sigma_{k+1}$$

(Proof)
$$\|A - A_k\|_2 = \left\| \sum_{j=k+1}^{p} \sigma_j u_j v_j^T \right\|_2$$
$$= \left\| U \begin{bmatrix} 0 & & & & & 0 \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \sigma_{k+1} & & \\ & 0 & & & \ddots & \\ & & & & & \sigma_p \\ \hline & & & 0 & & \end{bmatrix} V^T \right\|_2$$
$$= \left\| \begin{bmatrix} 0 & & & & & 0 \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \sigma_{k+1} & & \\ & 0 & & & \ddots & \\ & & & & & \sigma_p \\ \hline & & & 0 & & \end{bmatrix} \right\|_2$$ since U, V: orthogonal!

$= \sigma_{k+1}$ by definition of the matrix norm

Prove this as an exercise! ⇒ **Note**: If $D = \mathrm{diag}(d_1, \cdots, d_m) = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_m \end{bmatrix}$

then $\|D\|_p = \max_{1 \le j \le m} |d_j| \quad \forall p \ge 1$

Now, let $B \in \mathbb{R}^{m\times n}$ be any rank k matrix. Then $\dim(\mathrm{null}(B)) = n - k$

why? Because of the following Thm:

For any $A \in \mathbb{R}^{m\times n}$, $\mathrm{rank}(A) + \dim(\mathrm{null}(A)) = n$

Let $W := \text{null}(B) \cap \langle v_1, \cdots, v_{k+1} \rangle$

We know $W \neq \{0\}$ because

$\dim(\text{null}(B)) = n-k$

$\dim(\langle v_1, \cdots, v_{k+1} \rangle) = k+1$

so, if these two do not intersect, $\mathbb{R}^n$'s dimension would become $n-k+k+1 = n+1$

This cannot happen! #

So let $h \in W$, $h \neq 0$.

We can always normalize h, so can assume $\|h\|_2 = 1$.

Then,

$$
\begin{aligned}
\|A - B\|_2 &\geq \|(A-B)h\|_2 \quad \text{by def.} \\
&= \|Ah\|_2 \quad \text{since } h \in \text{null}(B) \\
&= \|U \Sigma V^T h\|_2 \\
&= \|\Sigma V^T h\|_2 \quad \text{since } U: \text{ortho.} \\
&\geq \sigma_{k+1} \|V^T h\|_2 \\
&= \sigma_{k+1} \|h\|_2 = \sigma_{k+1}
\end{aligned}
$$

since $h \in \langle v_1, \cdots v_{k+1} \rangle$

$$
V^T h = \begin{bmatrix} * \\ \vdots \\ * \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix} * \\ \vdots \\ * \end{matrix}}\right\} k+1 \\ \left.\vphantom{\begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}}\right\} n-k-1 \end{array}
$$

///

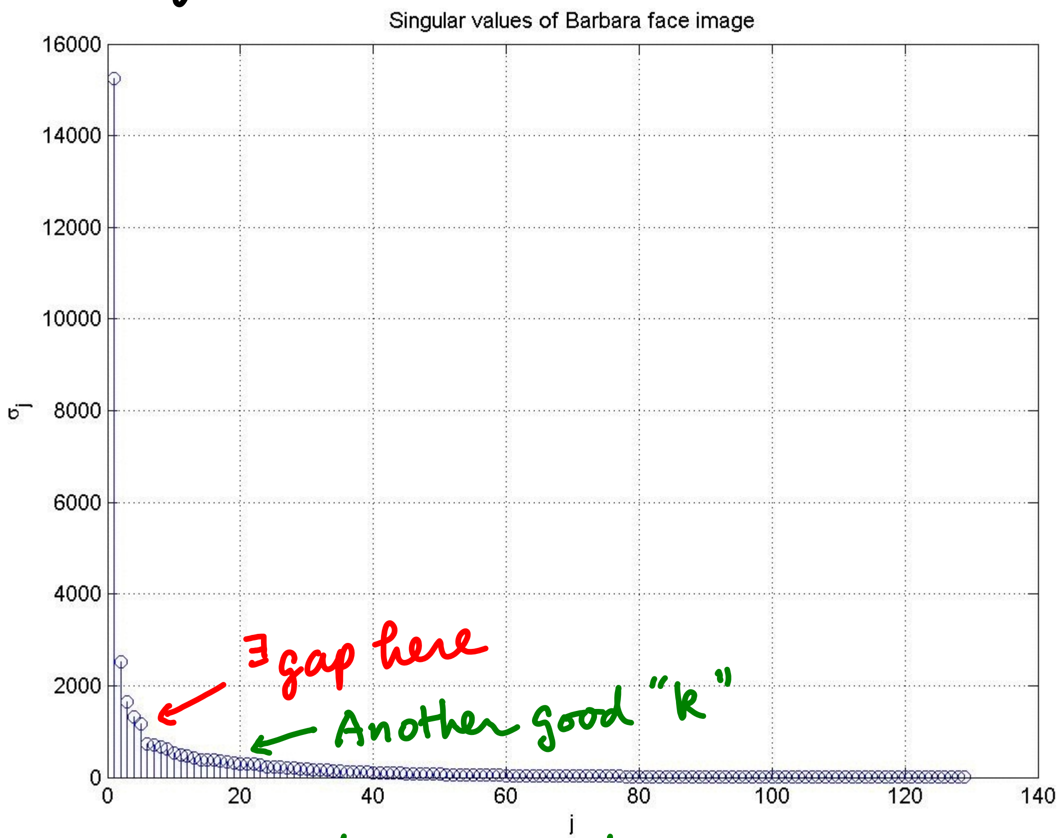
<u>Thm</u> For any k with $0 \leq k \leq r$,

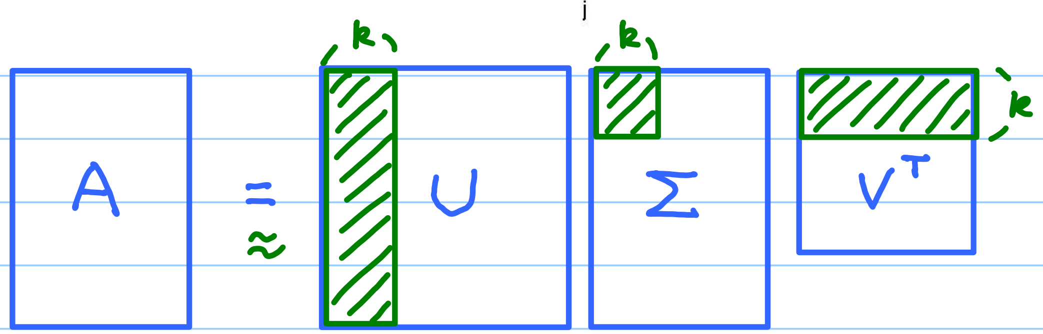
$$
\|A - A_k\|_F = \inf_{\substack{B \in \mathbb{R}^{m \times n} \\ \text{rank}(B) \leq k}} \|A - B\|_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}
$$

(Proof) Exercise!

So, for a given matrix, say, A how to determine a good "k" so that we can efficiently (i.e., compress) A without losing too much info of A?

$\Rightarrow$ Check the distribution of the singular values!





rank k approximation of A only uses ///// portions!

☆ Condition Number and SVD

Recall the condition number for a square nonsingular matrix A:

$$\kappa(A) = \text{cond}(A) := \|A\|_2 \cdot \|A^{-1}\|_2$$

$\kappa(A)$: small $\Rightarrow$ A: well-conditioned.
$\kappa(A)$: large $\Rightarrow$ A: ill-conditioned,
lose $\approx \log_{10}\kappa(A)$ digits to solve $A\mathbf{x} = \mathbf{b}$.

If A: singular, $\kappa(A) = +\infty$.

Using SVD of A, we can nicely compute $\kappa(A)$ as follows.

$\|A\|_2 = \sigma_1$ $\rightarrow$ by definition

$\|A^{-1}\|_2 = 1/\sigma_m$ $\xrightarrow{\text{why?}}$ $A^{-1} = (U\Sigma V^T)^{-1} = V\Sigma^{-1}U^T$
$= V \,\text{diag}(1/\sigma_1, \cdots, \underbrace{1/\sigma_m}_{\text{largest}})\, U^T$

So, $\kappa(A) = \sigma_1/\sigma_m$

We can generalize the definition of the condition number for a rectangular matrix $A \in \mathbb{R}^{m\times n}$ using the pseudo-inverse $A^\dagger$ and SVDs as

$$\kappa(A) := \|A\|_2 \cdot \|A^\dagger\|_2$$
$$= \sigma_1/\sigma_r \qquad r = \text{rank}(A) \le \min(m,n)$$